Assignment 4 - CIVENG C263H
Erin Jones
Due: 11/10/2023 (received extension due to walking pneumonia)

PART I

Question 1 - Report the total number of outflows and inflows

These should match, which they do, and the number 34,116,820 does match between inflows and outflows + is reported below

Question 2 - What is the total population of the given set of counties?

The total population, printed below, is 281,421,759

Question 3 - Ratio of Outflows:P

Given the fact that the ratio is small (much less than one), we can assume that there is not one outflow trip associated with every individual counted in the population. In fact, outflow trips only match ~12.12% of the population which can be attributed to many things. Firstly, for large counties, there may be a significant amount of self-looping where individuals are not commuting or leaving their counties much. In recent years, this has become even more common with the advent of teleworking. Furthermore, depending on how trips are calculated in this dataset, it could be that a single trip actually carries multiple individuals from one location to another (carpooling etc). Finally, if you look at a population pyramid for the US (e.g. https://www.populationpyramid.net/united-states-of-america/2020/) only ~63% is of working age and likely to commute across county lines.

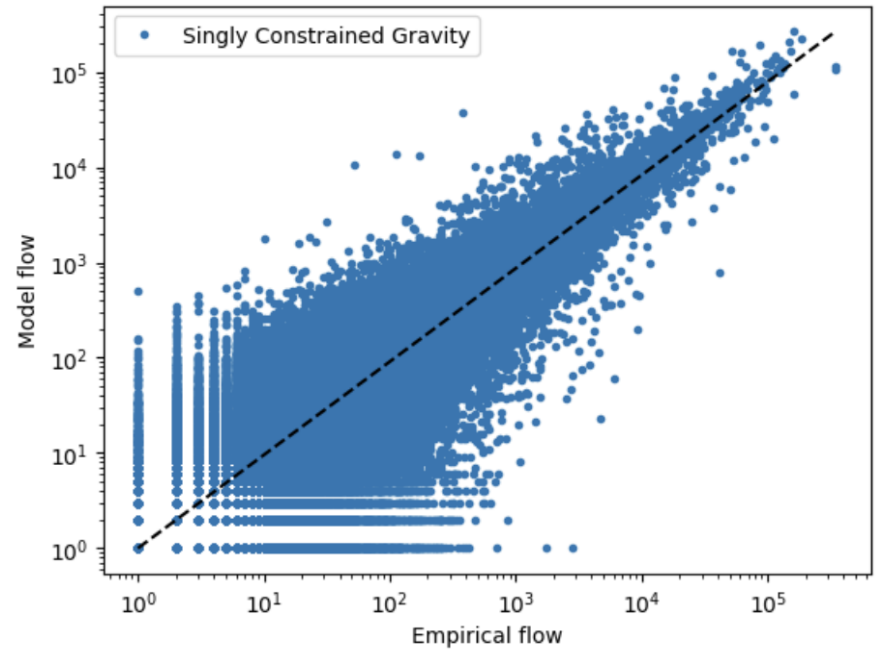
** UPDATE -- After further investivation, the outbound trips are based on the links file which contains ~3000 county IDs that are not in the nodes file which contains population information. This would lead to a significantly deflated outflows to pupulation ratio, especially depending on how many trips flow from the missing nodes.

Question 4 - Singly Constrained Gravitational Model

Generate the output using an outflow singly constrained gravity model with an exponential decaying function as the deterrence function type. Report the parameters of the model and show the scatter plot of the empirical flows based on the data vs. the estimates of synthetic flows created using the model. Select the Beta, as described in the Lenormand, et al.

Parameters

```
Gravity(name="Gravity model", deterrence_func_type="exponential", deterrence_func_args=[0.07882549547783631], origin_exp=1.0, destination_exp=1.0, gravity_type="singly constrained")
```

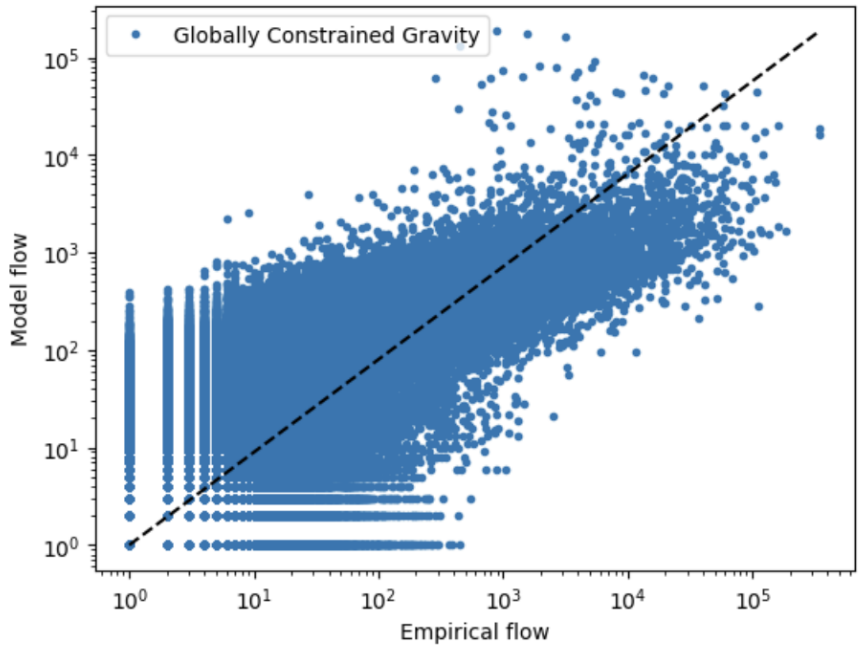


Question 5 - Globally Constrained Gravitational Model

Generate the output with globally constrained gravity model with a power law decaying function as the deterrence function type. Use the parameters reported by Viboud, et al.in Table 2 for trips larger than 119 km. Report the parameters of the model and show the scatter plot of data vs. the estimates of synthetic flows created using the model.

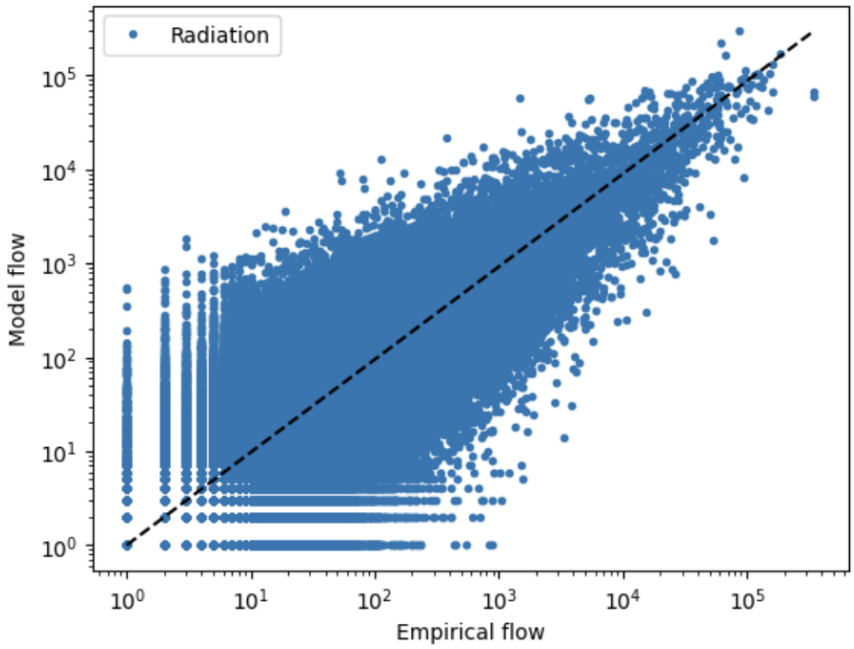
Parameters

```
Gravity(name="Gravity model", deterrence_func_type="power_law", deterrence_func_args=[-2], origin_exp=0.24, destination_exp=0.14, gravity_type="globally constrained")
```



Question 6 - Radiation Model

Apply the radiation model. Compare the data with the results of the radiation model by showing the scatter plot of the empirical flows based on the data vs. the estimates of synthetic flows created using the model.



Question 7 - Error Metrics

Report the common part of commuters (CPC), the root mean square error (RMSE), R-squared value, and maximum error of each model. How do the performances of these models compare with the reported values in the paper by Lenormand, et al. for the US commuting flow data?

```
The Metrics for the Singly Constrained Gravity Model with Exponential Deterrence are:
CPC = 0.7548147842006016
RMSE = 1376.0591357044593
R2 = 0.7369354995508939
Max Error = 233443

The Metrics for the Globally Constrained Gravity Model with Power-Function Deterrence are:
CPC = 0.3686699641638587
RMSE = 2650.7638571112275
R2 = 0.04193684174191507
Max Error = 330097

The Metrics for the Radiation Model are:
CPC = 0.6371990502563036
RMSE = 2251.2331524647625
R2 = 0.5645079255402387
Max Error = 279995
```

In our analysis, our **singly constrained gravity mode**l with exponential deterrence can be likened to the production constrained model using the gravity law with exponential function, non-normalized, because we did not normalize our results (and scikit-mobility's documentation does not indicate normalization) and the outflows are fixed for each location based on the data from links. In Figure 3, Lenormand and colleagues indicate that the CPC of their simulation is somewhere between 0.7 and 0.8, but closer to 0.8 - so I believe the result is in alignment (~0.7581).

Figure 3 reports the CPC of the **globally constrained gravity model** with power law deterrence to be slightly greater than 0.8 for the US - this result is very different from what we see above in the reported results for CPC. It seems, based on the R2 value, the model is not really fitting the data well, especially when compared to the other two models, which make me wonder whether some hyperparameter tuning/further calibration with the data might be needed for this dataset.

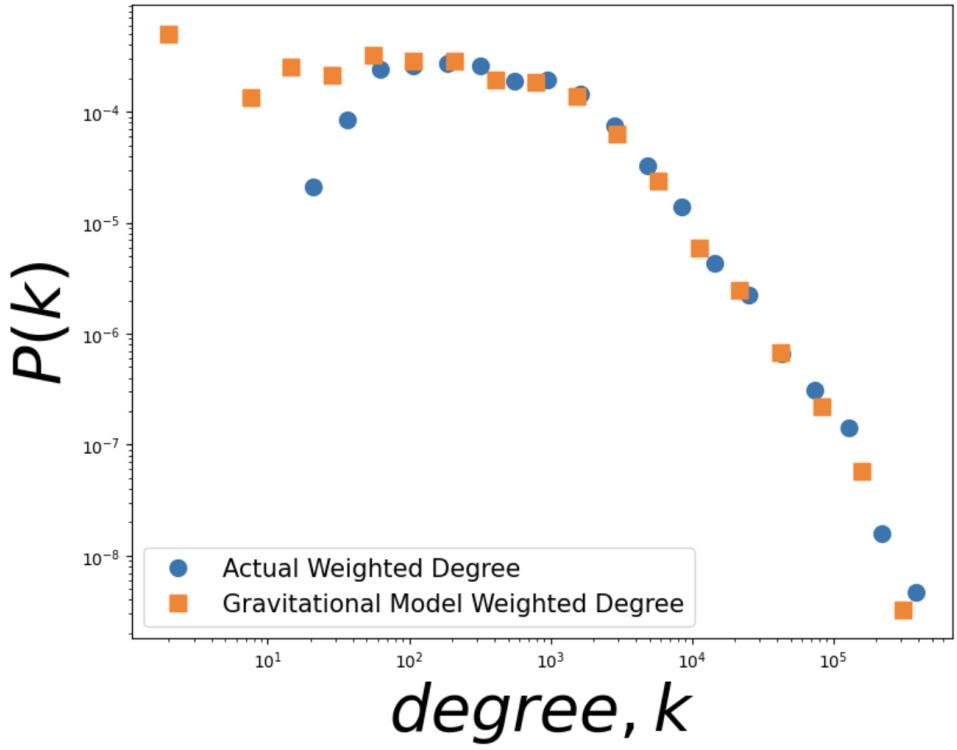
Figure 3 reports the CPC of the **radiation model** to hover between .6 and .7 in both partially and doubly constrained models with the extended model, which based on the scikit-mobility model is what the radiation model runs. This is in line with what we see in the results above. The model is fitting the data somewhere down the middle based on the r-squared metric reported above at 0.56.

In terms of the error metrics, Lenormand and colleagues report to calculating the RMSE, however they don't report on it in the published work, but turn instead to a mean absolute percent error. This value is generally going to be very different from the RMSE in cases where there are a lot of outliers, as MAPE punishes outliers less than RMSE. This means the error metrics are likely quite different in the case of the globally constrained model, where the model is not fitting the data well and producing high residuals.

PART II

Question 1

I recreated the graph for the singly constrained gravity model with exponential deterrence, as it performed the best out of the models.



Question 2

Network	**K_{min}**	**K_{max}**	**<k>**	σ_k	**<CC>**	**<L>**
US commuting data	21	654161	11380.112	0.031	3.46490176e-05	5.716619
Model US commuting data	0	601400	10993.4815	0.04688	3.3933	0.0